SYSTEMATIC DATA FARMING – AN APPLICATION TO A MILITARY SCENARIO

Choo Chwee Seng, Principal Member of Technical Staff
Ng Ee Chong, Senior Member of Technical Staff
Ang Choon Kiat, Senior Member of Technical Staff
Chua Ching Lian*, Member of Technical Staff

Operations Research Laboratory
DSO National Laboratories
20 Science Park Drive
Singapore 118230
SINGAPORE

ABSTRACT

This paper highlights the challenges of conducting simulation based experiments and describes how we seek to overcome these challenges through our implementation of a meta-technique known as Systematic Data Farming (SDF). We also describe its application on a military (Army) scenario to illustrate how the SDF capability can be used from the design phase, to the conduct phase and to the analysis phase. Through this application, we demonstrated the importance and value that the SDF capability can bring to simulation experiments. The paper will provide a detailed description of the process as well as the findings from the military scenario.

1. INTRODUCTION

Modelling, Simulation and Analysis (MS&A) plays an important role in our military's decision support framework, especially in the area of Experimentation and Operations Analysis (OA). As our Singapore Armed Forces (SAF) continues in the transformation towards a 3G SAF, there is an increasing need to conduct simulation-based experiments and studies that help explore new concepts of operations, investigate more scenarios, understanding the potential outcomes and capturing the surprises.

1.1 Key Challenges

For experiments and studies that are conducted for discovery purposes and are exploratory in nature, it is desirable to explore as many factors as possible and vary these factors over a wide range of levels or values. However, these requirements pose several challenges to conventional MS&A capabilities.

Classical Experiment Designs such as Factorial Designs become inefficient and even inadequate when the number of experimental factors and levels grow too large. For example, a Full Factorial design of 20 factors at 10 levels each will result in 10^{20} design points, which are almost intractable. Furthermore, the large amount of data generated makes analysis difficult, especially when the number of data points exceeded the input limitations of the analysis software.

Therefore, the current MS&A capability must be extended to provide a powerful, systematic and efficient approach to overcome these challenges.

1.2 Inspiration and Collaboration

The inspiration to develop the SDF capability is drawn from the work of Project Albert and our collaborators at the Naval Postgraduate School (NPS). Our collaboration with Project Albert helped established our principal expertise to set-up the data farming environment in DSO. We also worked closely with NPS to develop the knowledge of Advance Experiment Designs in the area of Latin Hypercube Designs.

2. THE SYSTEMATIC DATA FARMING PROCESS

The development of the SDF capability involved both collaboration and R&D work in the following 3 main areas: Data Farming, Advance Experiment Designs, and Clustering & Outlier Analysis.

2.1 Data Farming

Data Farming is a methodology developed by Project Albert that involves the use of high performance computer or computing grid to run a simulation thousands or millions of times across a large parameter and value space (Brandstein and Horne, 1998). Our collaboration with Project Albert experts involved setting up a Data Farming environment consisting of 32-CPUs within DSO that supports data farming requirements from both DSO projects and all other Project Albert collaborators. Our R&D work includes making non-agent based models data-farmable in our Data Farming environment.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **01 NOV 2006**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE **Systematic Data Farming An Application To A Military Scenario**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Operations Research Laboratory DSO National Laboratories 20 Science Park Drive Singapore 118230 SINGAPORE**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM002075., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**8**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

2.2 Advance Experiment Designs

All experiments should include some form of experimental design. Discovery experiments exploring new concepts can involve a large number of experimental factors, especially at the initial stage when the problem is still very open.

In this case, conventional factorial designs may not be practical as the resulting number of experiments will be too computationally expensive and time consuming. For an experiment involving 20 factors with a 10-level per factor set-up, the number of experiments to conduct based on a full factorial design is 10^{20}!

To overcome these problems faced, R&D work done by NPS recommends using statistical search methods to identify a set of good experimental design points (Kleijnen, Sanchez, Lucas and Cioppa, 2004). This resulting Advance Experiment Design is termed the Latin Hypercube (LHC) design and has markedly reduced the number of runs, hence helping to maximise information gained from the experiment when faced with constraints of time and resources. LHC designs have good space filling properties that reduce biasness and can be made nearly orthogonal for statistical efficiency (Ye, 1998; Cioppa, 2003). The trade-off for the reduced number of runs is that it only allows the main effects and some 2-factor interactions to be studied. However, this is usually sufficient for discovery experiments (Lucas et al, 2002).

Our collaboration with NPS involves using these LHC designs and extending this method to form Hybrid LHC Designs with Classical Factorial Designs or other customised designs. We also developed a Hybrid LHC generator to help generate these hybrid designs.

2.3 Clustering and Outlier Analysis

R&D was carried out on various powerful data-mining methods known to be capable of organising and analysing large quantities of data with the aim of identifying Clusters and Outliers.

K-Means methodology was coupled with Self-Organising Maps (SOM) to help organise the data into clusters. The incorporation of K-means was to help improve the clustering and segregation capability of the SOM (Vesanto and Alhoniemi, 2000).

Based on the Clusters identified, a search was carried out within to identify the points that are "most different" from the rest of the data points within the same cluster, ie. the outliers. This was achieved by comparing the Euclidean Distance of each data point with its k-nearest neighbour in each cluster and finding the one with the largest Euclidean Distance (Ramaswamy et al, 2000).

The result of our R&D effort was the use of hybrid clustering analysis techniques (k-means on self-organising maps) and outlier analysis to organise and extract "interesting" points or surprises from the large number of data points in the experiment. An analysis tool known as the Clustering and Outlier Analysis Data-Mining tool (COADM) was developed (Vesanto et al, SOM Toolbox for MATLAB).

2.4 Systematic Data Farming as a Process

Although the 3 components of the Systematic Data Farming (SDF) Capability are all useful tools on their own, we emphasize that they should be employed as an entire experimental and analytical process in experiments and studies. As illustrated in Figure 1, the proposed SDF process should involve the following steps:

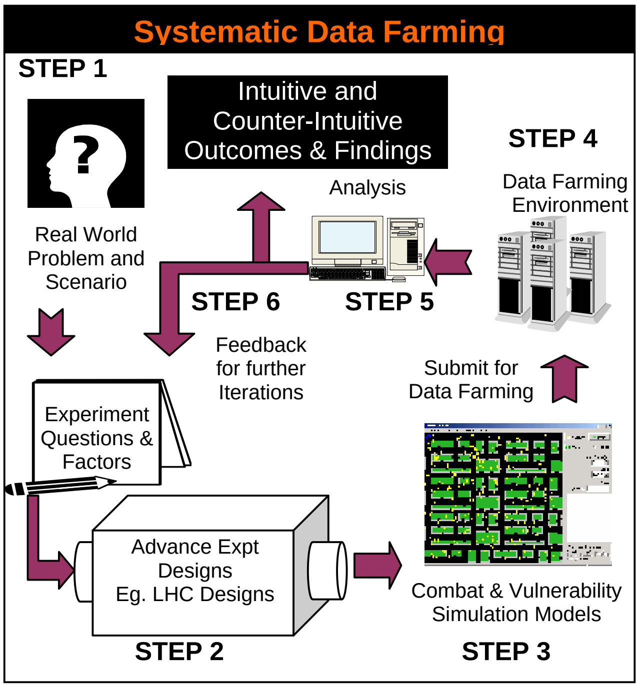


Figure 1 – The Systematic Data Farming Process

Step 1 - Scenario Specification. An appropriate vignette or scenario should be identified to scope the problem in the experiment or study.

Step 2 - Design of Experiment. Based on the questions to be identified in the experiment, a list of factors, each with the relevant range of levels, would be short-listed to be studied. The type of experiment design deemed suitable for the desired resolution and conduct of the experiment would be chosen, eg. LHC designs.

Step 3 - Simulation Models. A Simulation Model would be created to capture the important aspects of the scenario, especially those that are short-listed as factors in the experimental design. To fit into the SDF process, the model should be data-farmable using the data farming environment in DSO.

Step 4 - Data Farming. The simulation model and the experiment design are submitted for data farming using the data-farming environment. The results would be collected for analysis.

Step 5 – Regression and Clustering & Outlier Analysis. The analysis of the results should involve the co-operative use of statistical tools and the COADM tool to visualize and make sense of the results. The COADM tool should be applied to the data sets to provide a good overview of the output landscapes and relationships, highlighting the more influential factors and the clustering of design points. Analysis of outlier cases in the data set can be performed using the COADM tool. At the same time, statistical analysis can be conducted to examine these factors and identify the significant effects and interactions between the factors.

Step 6 - End of Process or Conduct Further Iterations. If the results have met the objectives of the experiment, the process can be terminated. Otherwise, the analyst should revisit the steps, do necessary modifications and perform further iterations to obtain more results.

3. APPLICATION OF THE SDF PROCESS

The rest of the paper describes the application of **ONE** iteration of the SDF process on an Army scenario and highlights the findings generated. Through this application, we seek to demonstrate how the challenges indicated under Section 1 were alleviated and illustrate the value that SDF can bring to simulation experiments.

4. THE SCENARIO

The Army scenario to be investigated pertains to an Urban Operation involving the raiding and capturing of a deliberately-defended Enemy Key Installation amidst the presence of hostile Civilians. Besides studying the contribution of platforms, sensors and weapon systems, the focus was to explore how the various intangible characteristics of the Blue Force, Red Force and Civilians affect the outcome of the operation. Examples of questions asked in the experiment include:

- How would Squad Cohesiveness and Aggressiveness affect the effectiveness and survivability of the Blue Force and Red Force?
- What would be the impact of Civilian behaviour on the Blue Force and Red Force effectiveness?

5. AGENT BASED SIMULATION MODEL

Based on the scenario described in Section 4, an Agent-Based Model was constructed using MANA. MANA, which stands for "**M**ap **A**ware **N**on-uniform **A**utomata", is an agent-based simulation tool developed by Defence Technology Agency, New Zealand. This tool was chosen because it has features that can represent both system-based and behavioral aspects of fighting forces. It was also a data-farmable and fast running tool, making it suitable for the SDF process.

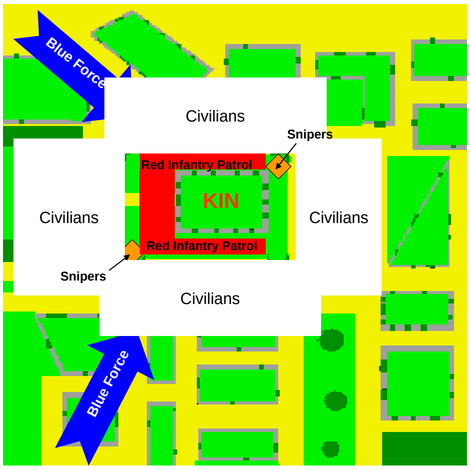


Figure 2 – Urban Scenario Setup in MANA

5.1 Scenario Set-up

An Urban Area of Operations 2km by 2km in size was set up in MANA. The scenario was set in this Urban AO where 2 platoons of Blue Infantry soldiers (21 soldiers per platoon), each platoon was supported by 3 MG-mounted soft-skin vehicles, attempted to take over a Key Installation (KIN) held by a platoon of Red Infantry soldiers (21 soldiers). The Red Infantry defence was assisted by two teams of Red snipers (4 snipers in total). The Blue agents' task was made more difficult by the crowd of hostile Civilians congregating near to the KIN and randomly attacking the Blue agents when they were encountered. The scenario setup is illustrated in Figure 2.

5.2 Modeling the Properties of Blue and Red Forces

Red and Blue Infantry agents were modelled slightly differently. The Blue Infantry agents were more mobile and were focused on reaching the objective, i.e. the KIN. The Red Infantry agents were more static and occupied defence positions around the KIN. The Blue Infantry agents had a higher probability to kill at shorter range and a higher rate of fire. The Red Infantry agents were given higher concealment rates, as they were considered to be more familiar with their environment. The Red sniper agents were given higher sensor range and probability to kill to reflect their enhanced sighting capability and longer range weapons.

Furthermore, the Red agents were hidden within the compounds of the building under cover and concealment, and the Red snipers were located within bunkers around the defending site. The Blue MG-mounted soft-skinned vehicles supporting the Blue Infantry agents were given higher protection and require greater number of hits to kill. Furthermore, their weapons were accorded higher probability to kill simulating the higher lethality of the machine guns.

5.3 Modeling of Civilians

The Civilians agents were dispersed within the AO around the KIN, and they had the tendency to congregate at the KIN, especially when Blue attacked the KIN. They were also naturally hostile to Blue agents and would attack Blue upon contact, although the civilians were configured to have low lethality. Their hostilities and behaviours towards the Blue agents were subjected to investigation in this study. Blue's Rule Of Engagement (ROE) against hostile Civilians would be to fire back only when attacked.

5.4 Modeling of Blue and Red Courses of Action

Apart from behaviour parameters, different Blue and Red courses of action were also modeled. There were 3 possible courses of action for the Blue Force and 2 for the Red Force. Blue Own Courses of Action (OCAs) are labelled OCA 1, OCA 2, & OCA 3 while Red Enemy Courses of Action (ECAs) are labelled ECA 1 & ECA 2. These are described as follows:

OCA 1. The Blue agents advanced from the northwest and southwest direction of the map towards the objective, attempting to take out the Red from both sides (see Figure 3 Blue arrows labeled "Blue OCA 1").

OCA 2. The Blue agents were concentrated in the southeast area of the map and advance as a force towards the Red, attempting to punch through the Red defence from a single direction (see Figure 3 Blue arrow labeled "Blue OCA 2").

OCA 3. The Blue agents were spread out on the northern portion of the map and attempted to flush out the read through a swarming approach (see Figure 3 Blue arrow labeled "Blue OCA 3").

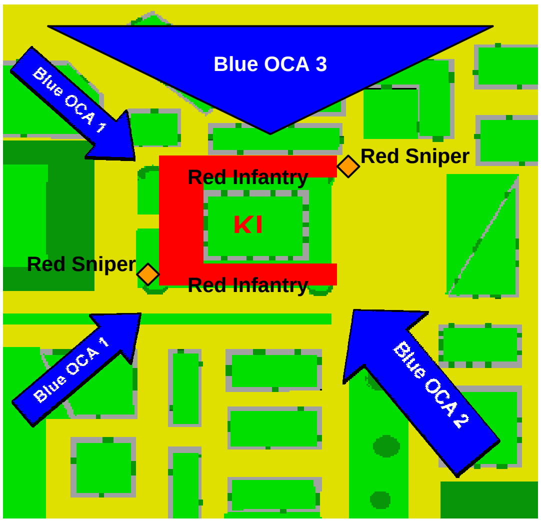
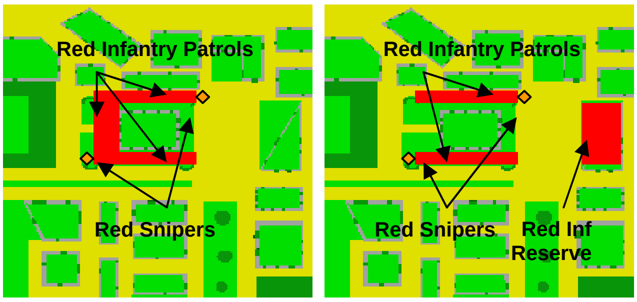


Figure 3 – Blue courses of action, OCA 1, 2 & 3.

ECA 1 - All Red agents resided within the building's compound and defended their base from there (see Figure 4 – Red ECA 1).

ECA 2. A section minus of 6 Red agents lay hidden in an adjacent building as backup to the other two sections in the defended locality. They were called in when the Red agents came in contact with Blue Forces (see Figure 4 – Red ECA 2).



Red ECA 1 **Red ECA 2**

Figure 4 – Red courses of action, ECA 1& 2.

6. DESIGN OF EXPERIMENT

To systematically study the scenario and derive useful analysis, a good experimental design is necessary.

6.1 Categorical Factors

The different Blue and Red courses of action were included in the design as 2 categorical factors, namely OCA and ECA. The full factorial design for these two categorical factors is as shown in Figure 5.

Design Points	OCA	ECA
Design Point 1	OCA 1	ECA 1
Design Point 2	OCA 1	ECA 2
Design Point 3	OCA 2	ECA 1
Design Point 4	OCA 2	ECA 2
Design Point 5	OCA 3	ECA 1
Design Point 6	OCA 3	ECA 2
Total 6 design Points		

Figure 5 – Full factorial design for OCA and ECA factors

6.2 Parametric Factors

As the experiment is exploratory in nature, a large sample space of the potential outcomes should be explored. A good way to do this would be to data farm the scenario over a large number of factors, each factor varied at fine resolution over a wide range of values.

A list of 30 parameters in the MANA scenario was short listed for data farming, with each parameter varied at 100 different levels within the Min and Max levels, as shown in Figure 6. As one of the focus of this study was to explore how the various intangible characteristics of the Blue Force, Red Force and Civilians would affect the outcome of the operation, the majority of the parameters short listed would affect the behaviour of the Blue, Red and Civilian agents in MANA.

6.3 Using the Latin Hypercube Experiment Design

Based on conventional Factorial Design, a 30-factor 100-level full factorial design would result in 100^{30} = $1x10^{60}$ design points! This is definitely too computationally and analytically intractable. A reduction in the resolution to vary the factors at only 20 levels each would still result in 20^{30} = $1x10^{39}$ design points, which is still computationally and analytically intractable.

Using the Latin Hypercube Generator developed under the SDF capability, a 30-factor 100-level Latin Hypercube (LHC) was generated. This LHC had 1000 design points, was nearly orthogonal at maximum correlation of 0.067, and had sufficient design points to study 2-factor interaction effects in a regression analysis.

6.4 Hybrid Latin Hypercube Experiment Design

To combine the 2 categorical factors design and the 30 parametric factors LHC design, a hybrid design was formed using the LHC Generator by crossing the 30-factor LHC with the 2-factor Full Factorial design for the OCA and ECA factors. The resultant hybrid design had 6000 design points and would be used in this study.

6.5 Measurements of Effectiveness

For the purpose of this exploratory study, the Measures of Effectiveness (MOEs) to be collected for analysis were:

- Total Blue Attrition.
- Total Red Attrition.
- Total Civilian Attrition.

Blue Inf Parameters	Min	Max
Cover And Concealment Level	-100	100
Tendency to Charge at KIN	-100	100
Tendency to Cluster with fellow Inf	-100	100
Individual Aggression Level	-100	100
Tendency to Move In Line Formation	-100	100
Squad Aggression Level	-100	100
Squad Cohesiveness Level	-100	100
Sensor Range	50	100
Mobility	50	200
Stealthiness	0	70
Blue Veh Parameters	**Min**	**Max**
Tendency to Move With Inf	-100	100
Tendency to charge at Enemy Inf	-100	100
Tendency to Fire At Snipers	-100	100
Tendency to charge at KIN	-100	100
Tendency to provide Inf Fire Support	-100	100
Sensor Range	50	100
Mobility	100	400
Red Inf Parameters	**Min**	**Max**
Cover And Concealment Level	-100	100
Tendency to Cluster with fellow Inf	-100	100
Tendency to Stay within KIN	-100	100
Individual Aggression Level	-100	100
Squad Aggression Level	-100	100
Squad Cohesiveness Level	-100	100
Stealthiness	0	70
Civilians Parameters	**Min**	**Max**
Initial Hostility against Blue	-100	100
Hostility after Contact with Blue	-100	100
Tendency to Cluster with fellow Civ	-100	100
Tendency to Cluster with fellow Civ After Contact wit Blue	-100	100
Tendency to Congregate at KIN	-100	100
Tendency to Congregate at KIN After Contact wit Blue	-100	100

Figure 6 – List of Parameters for Data Farming

7. DATA FARMING

The MANA model and the experimental design were submitted for data farming using the data farming facility in DSO. Based on the Hybrid LHC design, 6000 scenario excursions were generated from the 6000 design points. As the MANA model was stochastic in nature, each excursion was replicated 100 times and the mean MOEs for each excursion were computed. This resulted in a total of 600,000 runs which require around 2206 CPU hours of execution time. A single CPU will take around 91 days or 3 months to complete this data farming job!

However, with the parallel processing capability offered by the data farming environment, which comprised of 8x Intel P4 workstations and 9x nodes (each with 2x Intel Xeon processors), it took approximately just 85 hours or 3½days for this job to be completed. The output data was stored in CSV format and can be easily post-processed using Excel. The post processing was necessary for generating the required MOEs.

8. ANALYSIS OF RESULTS

The MOEs were analysed using 2 main methods, namely Statistical Analysis and Clustering & Outlier Analysis. The Statistical Analysis involved the use of linear regression methods available in many commercial statistical tools to analyse the data. This is quite established and would not be discussed in detail in this paper. However its findings will be compared with those obtained from the Clustering & Outlier Analysis.

The Clustering & Outlier Analysis was conducted using the COADM tool developed under the SDF capability and the following sections provided a more detailed description of the analysis and insights obtained.

8.1 Analysis using COADM

The large dataset of MOEs obtained from the data-farming output was analyzed using COADM and some interesting insights were derived.

Figure 7 shows some of the selected component plots of the SOM clusters generated by the COADM. Similar distribution of colors on the component plots implies correlation. Hence correlation between the factors and the MOEs can be discovered. Factors found to be correlated to MOEs are also the main factors contributing to the MOEs.

8.2 Analysis of Categorical factors

Both the OCA and ECA factors were observed to be uncorrelated with the MOEs. The distribution patterns of the OCA and ECA factors (shown on Figure 7) were observed to be rather independent from the distribution patterns of the MOEs. Hence, varying the OCA and ECA would not contribute to significant changes to the MOEs.

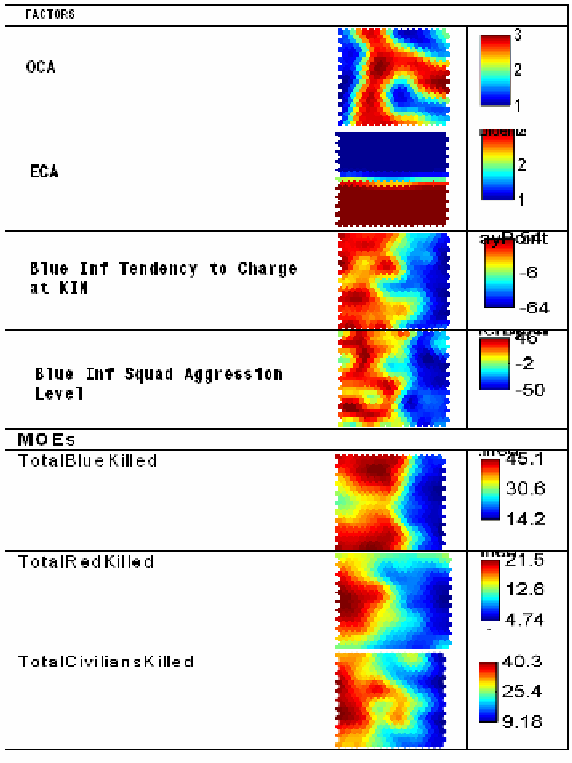


Figure 7- Component Plots of SOM clusters for selected Factors and MOEs.

8.3 Analysis of MOEs

The MOEs were observed to be somewhat correlated. This suggested that achieving high Red attrition would likely coincide with high Blue and Civilian attrition levels. The Red and Civilian casualties were more closely correlated with each other compared with that of the Blue casualties. Therefore, it would suggest that larger number of civilian casualties was unavoidable in this scenario, if the Blue agents or Red agents attempted to maximise the casualties on either sides.

However, there were exceptions. A region that contained outcomes that corresponded to moderate Blue attrition but very high Red attrition was shown in Figure 8. This would be the region of most interest to Blue as the parameter values defined in this region allowed Blue to achieve its mission of killing as many Red as possible without suffering high own attrition.

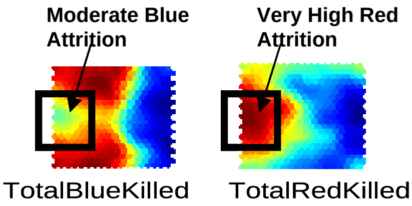
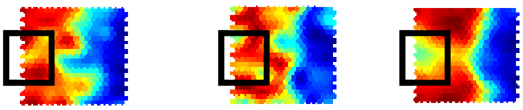


Figure 8- Region of Outcomes corresponding to Moderate Blue Attrition but Very High Red Attrition.

8.4 Analysis of Parametric Factors

Of the 32 farming parameters, it was observed that "Blue Infantry Tendency to Charge at KIN" and "Blue Infantry Squad Aggression Level" correlate most closely with the MOEs, and were hence most influential on the MOE outcomes.

It was interesting to revisit the region spotted under Figure 8, where Blue suffered moderate attrition but Red suffered high attrition. As shown in Figure 9, in this region, the parameter values for "Blue Infantry Tendency to Charge at KIN" and "Blue Infantry Aggression Level" should define the Blue's behavior that would inflict high Red attrition while sustaining moderate Blue attrition.



Blue Inf Tendency to Charge at KIN | **Blue Inf Squad Aggression Level** | **Total Blue Killed**

Figure 9 – Comparison of Blue Inf Tendency to Charge at KIN, Blue Inf Squad Aggression Level, and Total Blue Killed

8.5 Analysis of Clusters

COADM tool revealed that the data points can be organized into 20 clusters. The mean parameter values and MOEs for each cluster were obtained based on the data points within the cluster. By analyzing each cluster, we can identify the clusters that contained generally favorable outcomes for Blue and those that contained generally bad outcomes for Blue.

We can also identify contributing factors and behavior that resulted in each of these clusters. Without going into each cluster in detail, we would like to highlight that with this analysis, Blue would know how to manipulate Blue factors and make decisions to avoid those bad clusters and shift towards the good clusters.

8.6 Analysis of Outliers

From the output generated by COADM, the outlier points were examined in greater detail and they were laid out in Figure 10 in terms of the MOEs.

The top outlier was case number 5921 (or Data Point 5921) amongst the 6000 cases in the Experimental Design. This case belonged to Cluster 3 and had 23.45 Red killed in total. COADM identified this case as an outlier because 23.45 red killed was 1.936 times more than Cluster 3's mean value of total Red killed. A value that is 1.5 times either side of the mean would normally be considered as an outlier.

In Cluster 3, Blue generally suffers high attrition and hence Blue should avoid parameter values that will cause them to fall into this cluster. This outlier Case 5921 is an interesting case because it is the best outcome in a bad cluster for the Blue, as Blue was able to inflict much higher Red attrition compared to other cases in Cluster 3.

Case 5921 described a Blue force that was very fast, highly aggressive and extremely stealthy. Although the Red force and Civilians were also generally aggressive, they were less so compared to the Blue force.

Hence, if factors uncontrollable by the Blue Force, such as Red Force tactics and behavior, resulted in the circumstances becoming unfavourable (eg. falling into Cluster 3 outcomes), Blue force must attempt to exploit outlier case 5921 by moving swiftly and stealthily, and engaging more aggressively than the Red force inflict high Red casualties.

Case	Dist	Cluster	TotalBlueKilled	TotalRedKilled	TotalCiviliansKilled
5921	43.13	3	34.65 (+0.175)	23.45 (+1.936)	43.73 (+1.565)
4921	42.56	18	37.68 (+0.413)	22.88 (+1.838)	42.06 (+1.423)
1921	42.13	5	36.25 (+0.301)	23.63 (+1.966)	42.36 (+1.449)
921	41.93	11	37.89 (+0.430)	23.29 (+1.908)	40.92 (+1.327)
1115	41.31	5	40.47 (+0.633)	23.12 (+1.879)	46.93 (+1.835)
821	41.25	12	41.31 (+0.700)	21.83 (+1.657)	42.67 (+1.475)
2921	41.2	11	41.70 (+0.730)	20.24 (+1.385)	37.31 (+1.022)
1821	41.11	5	41.51 (+0.715)	20.69 (+1.462)	43.27 (+1.526)
3921	41.04	3	42.64 (+0.805)	20.34 (+1.402)	35.59 (+0.876)
762	40.99	12	29.84 (-0.205)	24.11 (+2.049)	45.98 (+1.755)

Figure 10 – MOEs in Outlier Cases.

8.7 Analysis & Findings from the Statistical Approach

The three MOEs, namely Total Blue Force attrition, Total Red Force attrition and Total Civilian attrition, were analysed separately using linear regression models that included main and two-factor interaction effects for the 32 factors (both categorical and parametric factors). This method provided information such as the statistical significance of the factors, the most influential factors, and the significant interactions between the factors.

The results showed that majority of the significant factors were Blue parameters. This implied that the Blue Force would be able to unilaterally affect the attrition levels of the Blue Force, Red Force and Civilians by employing the right set of behaviours and tactics.

It was also discovered through the analysis that the two most dominant factors that affected the MOEs were "Blue Infantry Tendency to Charge at KIN" and "Blue Infantry Squad Aggression Level". They dominated most interaction terms and more often than not, determined the contribution (+/-) of the interaction terms to the MOEs. This was consistent with the COADM analysis.

9. KEY ACHIEVEMENTS

An Army Scenario was modeled using Agent Based Simulation Models, where different behaviour and tactics for each type of agents were included.

A Hybrid LHC experiment design was generated to explore 2 categorical and 30 parametric factors. Such a design allowed the parametric factors to be studied over a large range of values and yet keep the total number of design points to just 6000, a manageable number.

The Hybrid LHC and the model were submitted to the Data Farming Environment for data farming, and the facility was able to handle and complete the 600,000 runs within 3½ days instead of weeks or even months.

The large dataset were analyzed using COADM and Statistical Analysis and the findings from both approaches showed good concurrence. The preliminary analysis performed produced interesting findings.

- The MOEs were highly correlated and hence high Red attrition would likely occur with high Blue and Civilian attrition, except for a specific identified region of parametric space that Blue can exploit.
- The OCAs & ECAs studied were unlikely to make much impact on the overall outcome.
- Certain Blue behaviour characteristics, such as aggression and tendency to charge at the KIN, were dominant factors. If these were manipulated correctly, Blue would likely be able to unilaterally improve their effectiveness in the operation.
- Outlier points showed that if Blue moved very swiftly & stealthily, and engaged Red more aggressively, it can still achieve a good outcome despite facing generally unfavourable conditions.

10. CONCLUSION

This paper briefly described the R&D work conducted on the SDF capability. We then focused on demonstrating the SDF capability employed in a military experiment based on an exploratory Urban Operations scenario. It was demonstrated that the SDF capability can overcome some of the key challenges of conducting a simulation experiment that seek to explore many factors and each factor varied at many levels. The paper concluded with a brief analysis of the rich landscape of outcomes obtained through the SDF process and the interesting findings were highlighted.

ACKNOWLEDGEMENTS

We would like to thank the following organizations and people that helped to make this R&D work possible:

Future Systems Directorate of the Singapore Armed Forces for funding this research.

Dr Gary Horne and Project Albert for sharing their expertise in Data Farming and Agent Based Modeling. Project Albert is an US Marine Corps' research effort to assess the general applicability of the study of Complex Adaptive Systems (CAS) to land warfare, and to provide new methodologies for investigating the results of running such models, and incorporating those results with other more traditional, methods of analysis.

Prof Thomas Lucas and Prof Susan Sanchez of NPS for sharing valuable knowledge and reference material on Latin Hypercube Designs.

Defence Technology Agency, New Zealand Defence Force, for sharing the Agent Based Model, MANA under the collaboration in Project Albert.

REFERENCES

Alfred Brandstein and Gary Horne, 1998: Data Farming: a Meta-Technique for Research in the 21st Century, Maneuver Warfare Science 1998, Marine Corps Combat Development Command Publication.

Jack P.C. Kleijnen, Susan M. Sanchez, Thomas W Lucas and Thomas M. Cioppa, 2004: A User's Guide to the Brave New World of Designing Simulation Experiments

J. Vesanto, and E. Alhoniemi, 2000: "Clustering of Self-Organising Map", IEEE Transactions on Neural Networks, volume 11, number 3, pages 586.

J. Vesanto, E. Alhoniemi, K. Kiviluoto, and J. Parvianen: Self- Organising Map for Data Mining in Matlab: SOM Toolbox. www.cis.hut.fi/projects/somtoolbox

S. Ramaswamy, R. Rastogi, and K. Shim. 2000: Efficient algorithms for mining outliers from large datasets. In Proceedings of the ACM SIGMOD Conference, pages 427–438.

Thomas M. Cioppa, 2003: Advanced Experimental Designs for Military Simulations, Technical Document TRAC-M-TR-03-11, 2003

Thomas W. Lucas, Susan M. Sanchez, Lloyd Brown and William Vinyard, 2002: Better Designs for High-Dimensional Explorations of Distillations, Maneuver Warfare Science 2002, Marine Corps Warfighting Lab Publication.

Ye K.Q, 1998: Orthogonal Column Latin Hypercubes and Their application in Computer Experiments, Journal of the American Statistical Association – theory and methods, Vol 93, No. 444, pp. 1430-1439.